PCT/GB 2003 / 0 0 3 4 6 5

INVESTOR IN PEOPLE

GB03/03465

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

REC'D 09 SEP 2003
WIPO PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 20 August 2003

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

1/77

THE PATENT OFFICE
A
19 AUG 2002
RECEIVED BY FAX

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

19AUG02 E742068-1 D02776
P01/7700 0.00-0219271.4

1. Your reference — IP/P3302/1

2. Patent application number *(The Patent Office will fill in this part)* — 0219271.4 — 19 AUG 2002

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

QINETIQ LIMITED

Registered Office 85 Buckingham Gate
London SW1E 6PD
United Kingdom

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation — GB

GB 8183854001

4. Title of the invention — Histological Assessment

5. Name of your agent *(if you have one)* — Bowdery Anthony Oliver

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

QINETIQ LIMITED
IP Formalities
A4 Bldg
Cody Technology Park
Ively Road
Farnborough
Hants GU14 0LX United Kingdom

8262792002

Patents ADP number *(if you know it)*

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*
GB	0218909.0	15/08/2002

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number or earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body. See note (d))

YES (b)

Patents Form 1/77

Patents Form 1/77

0044757 19 AUG 02 03:36

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description 23 ✓ Wa

Claim(s) 0

Abstract 0

Drawing(s) 5 only ✓

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*

Request for preliminary examination and search *(Patents Form 9/77)*

Request for substantive examination *(Patents Form 10/77)*

Any other documents *(please specify)*

11. I / We request the grant of a patent on the basis of this application.

Signature Dr A W S Williams Date 19.8.02

12. Name and daytime telephone number of person to contact in the United Kingdom Mrs Linda Bruckshaw 01252 392722

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent of the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have attached 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

Patents Form 1/77

Coloured Image Assessment

This invention relates to a method, an apparatus and a computer program for assessment of coloured images: it is particularly (although not exclusively) relevant to assessment of histological slides to provide clinical information on potentially cancerous tissue such as breast cancer tissue.

Breast cancer is a common form of female cancer: once a lesion indicative of breast cancer has been detected, tissue samples are taken and examined by a histopathologist to establish a diagnosis, prognosis and treatment plan. However, pathological analysis of tissue samples is a time consuming and inaccurate process. It entails interpretation of colour images by human eye, which is highly subjective: it is characterised by considerable inaccuracies in observations of the same samples by different observers and even by the same observer at different times. For example, two different observers assessing the same ten tissue samples may easily give different opinions for three of the slides - 30% error. The problem is exacerbated by heterogeneity, i.e. complexity of some tissue sample features. Moreover, there is a shortage of pathology staff.

There is a need to provide an objective measurement of Cerb-B2, ER and PR status and vascularity to contribute to an effective treatment plan for the patient.

In order that the invention might be more fully understood, embodiments thereof will now be described, by way of example only, with reference to the accompanying drawings, in which:-

Figure 1 is a block diagram of a procedure for measuring indications of cancer to assist in formulating diagnosis and treatment;

Figure 2 is a block diagram of a process for measuring Cerb-B2 in the procedure of Figure 1;

Figure 3 is a block diagram of a process for measuring vascularity in the procedure of Figure 1;

Figure 4 is a block diagram of a process for measuring oestrogen and progesterone receptor in the procedure of Figure 1;

Figure 5 is a pseudo three dimensional view of a red, green and blue colour space (colour cube) plotted on respective orthogonal axes;

Figure 6 is a transformation of Figure 5 to form a chromaticity space;

Figure 7 is a drawing of a chromaticity space reference system; and

Figure 8 illustrates use of polar co-ordinates.

The examples of invention to be described herein are three different inventions which can be implemented separately or together, because they are all measurements which individually or collectively assist a clinician to diagnose cancer and to formulate a treatment programme. In descending order of importance, the procedures are determination of oestrogen and progesterone receptor, determination of Cerb-B2 and determination of vascularity. For convenience however, they are described in the order determination of Cerb-B2, determination of vascularity and determination of oestrogen and progesterone receptor.

A procedure 10 for the assessment of tissue samples in the form of histopathological slides of potential carcinomas of the breast is shown in Figure 1. This drawing illustrates processes, which generate measurements of specialised kinds for use by a pathologist as the basis for assessing patient diagnosis, prognosis and treatment plan.

The procedure 10 employs a database, which maintains digitised image data obtained from histological slides as will be described later. Sections are taken (out) from breast tissue samples (biopsies) and placed on respective slides. Slides are stained using a staining agent selected from the following depending on which parameter is to be determined:

a) Immunohistochemical staining for Cerb-B2 with diaminobenzidine (DAB) as substrate (chemical staining agent) - collectively "Cerb-DAB" - this is for assessing Cerb-B2 gene amplification status;

b) Oestrogen receptor (ER) with DAB as substrate (collectively "ER-DAB") for assessing the expression (the amount expressed or emitted) of the oestrogen receptors. Progesterone receptor (PR) status is investigated using chemical treaqtment giving the same colouration as in ER.

c) Immunohistochemical staining for CD31 with fuchsin (F) as substrate for assessing vascularity (angiogenesis).

In a prior art manual procedure, a clinician places a slide under a microscope and examines a region of it (referred to as a tile) at magnification of x40 for indications of Cerb-B2, ER and PR status and vascularity.

The present invention requires data from histological slides in a suitable form. In the present example, image data were obtained by a pathologist using Zeiss Axioskop microscope with a Jenoptiks Progres 3012 digital camera. Image data from each slide is a set of digital images obtained at a linear magnification of 40 (i.e. 40X), each image being an electronic equivalent of a tile.

To select images, a pathologist scans the microscope over a slide, and at 40X magnification selects regions (tiles) of the slide which appear to be most promising in terms of an analysis to be performed. Each of these regions is then photographed using the microscope and digital camera referred to above, which produces for each region a respective digitised image in three colours, i.e. red, green and blue (R, G & B). Three intensity values are obtained for each pixel in a pixel array to provide an image as a combination of R, G and B image planes. This image data is stored temporarily at 12 for later use.

Two tiles are required for vascularity measurement at 14, and one tile for each of oestrogen and progesterone receptor measurement at 16 and Cerb-B2 measurement at 18. These measurements provide input to a diagnostic report at 20.

The prior art manual procedure for scoring Cerb-B2 involves a pathologist subjectively and separately estimating stain intensity, stain location and relative number of cells associated with a feature of interest in a tissue sample. The values obtained in this way are combined by a pathologist to give a single measurement for use in diagnosis, prognosis and reaching a decision on treatment. The process hereinafter described in this example replaces the prior art manual procedure with an objective procedure.

Referring now to Figure 2, a flow diagram of the Cerb-B2 measurement process 18 is shown in more detail: the process 18 is applied to one image or tile obtained by magnifying by a factor of 40 an area of a histological slide. The image is a three colour or RGB image as defined above, i.e. there is a respective image plane for each colour. For the purposes of the following analysis, the letters R, G and B for each pixel are treated as the red green and blue intensities at that pixel. The RGB input image is used

at 30 to compute a cyan image derived from the blue and green image planes: i.e. for each pixel a cyan intensity C is computed from $C = (2 \times B + G)/3$, the respective pixel's green (G) intensity being added to twice its blue (B) intensity and the resulting sum being divided by three. When repeated for all pixels this yields a cyan image or image plane.

At 32, a Sobel edge filter is applied to the cyan image plane: this is a standard image processing technique published in Klette R., & Zamperoni P., 'Handbook of image processing operators', John Wiley & Sons, 1995. A Sobel edge filter consists of two 3x3 arrays of numbers S_P and S_Q, each of which is convolved with successive 3x3 arrays of pixels in an image. Here

$$S_P = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \text{ and } S_Q = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \qquad (1)$$

The step 32 initially selects a first cyan 3x3 array of pixels in the top left hand corner of the cyan image: designating as C_{ij} a general cyan pixel in row i and column j, the top left hand corner of the image consists of pixels C_{11} to C_{13}, C_{21} to C_{23} and C_{31} to C_{33}. C_{ij} is then multiplied by the respective digit of S_P located in the S_P array as C_{ij} is in the 3x3 cyan pixel array: i.e. C_{11} to C_{13} are mutiplied by 1, 2 and 1 respectively, C_{21} to C_{23} by zeroes and C_{31} to C_{33} by -1, -2 and -1 respectively. The products so formed are added algebraicly and provide a value p.

The value of p will be relatively low for pixel values changing slowly between the first and third rows either side of the row of C_{22}, and relatively high for pixel values changing rapidly between those rows: in consequence p provides an indication of image edge sharpness across rows. This procedure is repeated using the same pixel array but with S_Q replacing S_P, and a value q is obtained: q is relatively low for pixel values changing slowly between the first and third columns either side of the column of C_{22}, and relatively high for pixel values changing rapidly between those columns: and q therefore provides an indication of image edge sharpness across columns. The square root of the sum of the squares of p and q are then computed i.e. $\sqrt{p^2+q^2}$, which is defined as an "edge magnitude" and becomes T_{22} (replacing pixel C_{22} at the centre of the 3X3 array) in the transformed cyan image. It is also possible to derive an edge "phase angle" as $\tan^{-1}p/q$, but that is not required in the present example.

A general pixel T_{ij} (row i, column j) in the transformed image is derived from $C_{i-1,j-1}$ to $C_{i-1,j+1}$, $C_{i,j-1}$ to $C_{i,j+1}$ and $C_{i+1,j-1}$ to $C_{i+1,j+1}$ of the cyan image. Because the central row and column of the Sobel filters in Equation (1) respectively are zeros, and other coefficients are 1s and 2s, p and q for T_{ij} can be calculated as follows:

$$p = \{ C_{i-1,j-1} + 2C_{i-1,j} + C_{i-1,j+1} \} - \{ C_{i+1,j-1} + 2 C_{i+1,j} + C_{i+1,j+1} \} \quad (2)$$

$$q = \{ C_{i-1,j-1} + 2C_{i,j-1} + C_{i+1,j-1} \} - \{ C_{i-1,j+1} + 2C_{i,j+1} + C_{i+1,j+1} \} \quad (3)$$

Beginning with $i = j = 2$, p and q are calculated for successive 3X3 pixel arrays by incrementing j by 1 and evaluating Equations (2) and (3) for each such array until the end of a row is reached; j is then incremented by 1 and the procedure is repeated for a second row and so on until the whole image has been transformed. This transformed image is referred to below as the "Sobel of Cyan" image or image plane.

The Sobel filter cannot calculate values for pixels at image edges having no adjacent pixels on one or other of its sides: i.e. in a pixel array having N rows and M columns, edge pixels are the top and bottom rows and the first and last columns, or in the transformed image pixels T_{11} to T_{1M}, T_{N1} to T_{NM}, T_{11} to T_{1M} and T_{1M} to T_{NM}. By convention in Sobel filtering these edge pixels are set to zero.

The next step 34 is to compute the mean and standard deviation of the transformed pixel values T_{ij}. For convenience a change of nomenclature is implemented: index k is substituted for i and j, i.e. k = 1 to NM for i, j = 1, 1 to N, M: this treats a two dimensional image as a single composite line composed of successive rows of the image. Also x is substituted for T in each pixel value, so T_{ij} becomes x_k. The following Equations (4) and (5) respectively are used for computing the mean μ and standard deviation σ of the transformed pixels x_k.

$$\mu = \frac{1}{NM} \sum_{k=1}^{NM} x_k \quad (4)$$

$$\sigma = \sqrt{\frac{1}{NM-1} \sum_{k=1}^{NM} (x_k - \mu)^2} \quad (5)$$

At 36, various statistical parameters are computed for the Red, Green, Blue and Cyan image planes using Equations (4) and (5) above and Equation (6) below.

$$Skewness = \frac{1}{NM}\sum_{k=1}^{NM}\left[\frac{x_k - \mu}{\sigma}\right]^3 \qquad (6)$$

For the Red image plane the statistical parameters are the mean μ, standard deviation σ and skewness of its pixel values: in Equations (4) to (6) x_k (k = 1 to NM) represents a general pixel value in the Red image plane. In addition, the Red image plane's pixels are compared with one another to obtain their maximum, minimum and range (maximum – minimum). Similarly, pixels in each of the Green and Blue image planes are compared with one another to obtain a respective maximum, minimum and range for each of these planes. Finally, for the Cyan image, pixels' mean and standard deviation are computed using Equations (4) and (5), in which x_k represents a general pixel value in the Cyan image plane.

At 38, a colour segmentation is computed from the Red, Green, Blue, Cyan and Sobel of Cyan image planes. Step 38 is shown in more detail within chain lines 40. Colour segmentation is implemented by applying three sets of adaptive thresholding operations referred to as D, E and F respectively. Each set of operations involves two respective numerical factors: D has factors D1 = 0.802 and D2 = 1.24, E has E1 = E2 = 0.903, and F has F1 = 1.24 and F2 = 0.802.

The first set of thresholding operations is D using factors D1 and D2. It implements the following steps:

(a) Produce a thresholded image for the Red image plane as follows: for every Red pixel value that is less than an adaptive threshold, set the corresponding pixel location in the thresholded image to 1, otherwise set the latter to 0. A respective adaptive threshold is computed separately for every pixel location as follows:

(i) Select the same pixel location in the cyan image and from it search in four directions - the north, south, east and west directions - for seventy pixel locations (or as many as are available up to seventy). Here north, south, east and west have the following meanings: north: upwards from the pixel in the same column; south: downwards from the pixel in the same column; east:

rightward from the pixel in the same row; and west: leftward from the pixel in the same row. More directions could be used to improve accuracy but four have been found to be adequate for the present process. If any of the seventy locations in a respective direction has a pixel value less than the product of D2 and the Cyan mean μ_C then the direction is assigned a value of 1, otherwise it is assigned a value of 0.

(ii) Sum all four direction values to provide a sum S_d and set the adaptive threshold for the current pixel location to whichever is the lesser of 255 and (D1 × Red mean μ_R) + (S_d - 4) x Red standard deviation σ_R).

(b) Produce a thresholded image for the Cyan image plane as follows: using the Cyan mean μ_C from step 34, for every Cyan pixel value that is less than (D2 x μ_C), set the pixel in the corresponding location in the thresholded image to 0, otherwise set the latter to 1. This has the effect of removing excess brown pixels.

(c) Produce a thresholded image for the transformed or Sobel of Cyan image plane as follows: using the Cyan mean μ_C and standard deviation σ_C from step 34: i.e. for every Cyan pixel value that is greater than (μ_C + 1.5σ_C) set the corresponding pixel in the thresholded image to 0, otherwise set the latter to 1. This has the effect of removing excess edge pixels.

(d) Using the pixel minimum and range values computed at step 36, a single thresholded binary image is now produced using data obtained from the Red, Green and Blue image planes. This is carried out as follows: for each Red, Green and Blue pixel group at a respective pixel location that satisfies all three criteria at (i) to (iii) below, set the pixel at the corresponding location in the thresholded image to 0, otherwise set the latter to 1; this has the effect of removing lipid image regions (regions of fat which appear as highly saturated white areas). The criteria for each co-located Red, Green and Blue pixel group are:

(i) Red pixel > Red minimum + 0.98 × (red range), AND

(ii) Green pixel > Green minimum + 0.98 × (green range), AND

(iii) Blue pixel > Blue minimum + 0.98 × (blue range)

(e) The next step is to apply to the binary image obtained at 38d above a series of morphological closing operations, which consist of a dilation operation followed by an erosion operation. These morphological operations are to fuse narrow gaps and eliminate small holes in individual groups of contiguous pixels appearing as blobs in an image. It can be thought of as removal of irregularities or spatial "noise", and it is a standard image processing procedure published in Umbaugh S.C., 'Colour vision and image processing', Prentice Hall, 1998. The number of operations applied depends on circumstances: operations (i.e. filling small gaps and/or holes) terminate when the number of changes made by the erosion operation is greater than the number made by the dilation operation.

(f) A connected component labelling process is now applied to the binary image produced at 38e. This is a known image processing technique (sometimes referred to as 'blob colouring') published by R Klette and P Zamperoniu, 'Handbook of Image Processing Operators', John Wiley & Sons, 1996, and A Rosenfeld and A C Kak, 'Digital Picture Processing', Vols. 1 & 2, Academic Press, New York, 1982. It gives numerical labels to "blobs" in the binary image, blobs being regions or groups of like-valued contiguous or connected pixels in an image: i.e. each group or blob consists of connected pixels which are all 1s, and each is assigned a number different to those of other groups. This enables individual blobs to be distinguished from others by means of their labels. The number of labelled image regions or blobs in the image is computed from the labels and output. Connected component labelling also determines each labelled image region's centroid (pixel location of region centre) and area, which are used later as will be described.

As indicated by arrows 42, the second and third sets of thresholding operations E and F are carried out by iterating steps 38a to 38f above using factors E1/E2 and F1/F2 respectively.

(g) A technique referred to as the Downhill Simplex method is now applied to the results of the thresholding operations obtained from 38a to 38f above. It is a standard iterative statistical technique for multidimensional optimisation published in Nelder J.A., Mead R., 1965, Computer Journal, vol. 7, pp 308-313, 1965. In the present example the technique is used to determine which pair of factors D1/D2, E1/E2 or F1/F2 used above yields the maximum number of image regions or

groups of connected pixels. It takes as input the three pair of factors D1/D2, E1/E2 and F1/F2 together with the function returning the number of detected regions and described at 38a and 38b above for adaptive thresholding.

The procedure is now concerned with determining quantities referred to as "grand mean" and "mean range" to be defined later. If the Downhill Simplex method 38g has determined that there are not more than a user specified number of image regions, sixteen in the present example, then at 44 processing is switched to 46: here both grand mean and mean range are set to 1.0 and processing passes to step 54 (to be described later) to compute a required result.

If the Downhill Simplex method has determined that there are more than sixteen image regions, then at 44 processing is switched to 48 where a search to characterise these regions' boundaries is carried out. The search uses each region's area and centroid pixel location as obtained in connected component labelling at 38f, and each region is assumed to be a cell with a centroid which is the centre of the cell's nucleus. This assumption is justified for most cells, but there may be misshapen cells for which it does not hold: it is possible to discard misshapen cells by eliminating those with concave boundary regions for example, but this is not implemented in the present example.

The search to characterise the regions' boundaries is carried out along the respective north, south, east and west directions (as defined earlier) from the centroid: it is carried out in each of these directions for a distance δ which is either 140 pixels or $2\sqrt{region\ area}$, whichever is the lesser. It employs the cyan image because experience shows that this image gives the best defined cell boundaries with the slide staining previously described. Designating C_{ij} as the intensity of a region's centroid pixel in the cyan image at row i and column j, then pixels to be searched north, south, east and west of this centroid will have intensities in the cyan image of $C_{i+1,j}$ to $C_{i+\delta,j}$, $C_{i-1,j}$ to $C_{i-\delta,j}$, $C_{i,j+1}$ to $C_{i,j+\delta}$ and $C_{i,j-1}$ to $C_{i,j-\delta}$ respectively. The cyan intensity of each of the pixels to be searched is subtracted from the centroid pixel's cyan intensity C_{ij} to produce a difference value, which may be positive or negative. In a cyan image, a cell nucleus is normally blue whereas a boundary is brown (with staining as described earlier).

Each pixel is then treated as being part of four linear groups (or windows) of six, twelve, twenty-four and forty-eight pixels each including the pixel and extending from it in a

continuous line north, south, east or west according respectively to whether the pixel is north, south, east or west of the centroid. $C_{i+1,j}$ for example is grouped with $C_{i+2,j}$ to $C_{i+6,j}$, $C_{i+2,j}$ to $C_{i+12,j}$, $C_{i+2,j}$ to $C_{i+24,j}$, and $C_{i+2,j}$ to $C_{i+48,j}$ (inclusive in each case). This provides a total of 16δ groups from 4δ groups in each of four directions. For each group the difference between each of its pixels' cyan intensities and that of the centroid is calculated: the differences are summed over the group algebraically (positive and negative differences cancelling one another) and divided by the number of pixels in the group to provide a net difference per pixel between the cyan intensities of the group's pixels and that of the centroid.

For each direction, i.e. north, south, east and west, there is now a respective set of 4δ net differences per pixel: in each set the net differences per pixel are compared and their maximum value is identified. This produces a respective maximum net difference per pixel for each of the sets, i.e. for each of the north, south, east and west directions, and size of window (number of pixels in group) in which the respective maximum occurred. The four maxima so obtained (one for each direction) and the respective window size in each case are stored. Each maximum is a measure of the region boundary (cell membrane) magnitude in the relevant direction, because in a cyan image the maximum difference as compared to a blue cell nucleus occurs at a brown boundary. The window size associated with each maximum indicates the region boundary width, because a boundary width will give a higher maximum in this technique with a window size which it more nearly matches as compared to one it matches less well. Greater accuracy is obtainable by using more window sizes. A further option is to record the position of the maximum or boundary as being that of one of the two pixels at the centre of the window in which the maximum occurs: this was not done in the present example, although it would enable misshapen cells to be detected and discarded as being indicated by significant differences in the positions of maxima in the four directions.

The next step 50 is to apply what is referred to as a "quicksort" to the four boundary magnitudes to sort them into ascending order. Quicksort is a known technique published by Klette R., Zamperoniu P., 'Handbook of Image Processing Operators', John Wiley & Sons, 1996, and will not be described. For each image region, measurements made as described above are now recorded in a respective 1-dimensional vector as set out in Table 1 below.

TABLE 1

Element number	Parameter
0	North boundary magnitude
1	South boundary magnitude
2	East boundary magnitude
3	West boundary magnitude
4	Region area (from 38f)
5	North boundary width
6	South boundary width
7	East boundary width
8	West boundary width
9	Sum of North, South, East and West boundary magnitudes

A further quicksort is now applied (also at 50) to the regions to sort them into ascending order of element 9 values in Table 1 above, i.e. sum of north, south, east and west boundary magnitudes. A subset of the regions is now selected as being those having large values of element 9; these most significant regions are typically the top one eighth of the subset of regions in terms of element 9 magnitude. From this subset of regions the following parameters are computed at 52, "grand mean", "mean range" and "relative range" as defined below :

octile = one eighth of the number of regions in the subset (8)

boundaries = boundary magnitudes (9)

Σ = sum of ... (over all regions in the subset) (10)

element 1= south boundary magnitude (11)

element 3 = west boundary magnitude (12)

grand mean = 6×[(Σ east boundaries) + (Σ west boundaries)
+ (Σ north boundaries) + (Σ south boundaries)]/ 4octile (13)

mean range = [(Σ boundary element 3) - (Σ boundary element 1)]/octile. (14)

relative range = 10xmean range/grand mean (15)

At 54 an overall distance measure is computed: this measure provides an estimate of how far the current cyan image (used to derive grand mean etc above) is from each member of a predetermined standard set of images, four images in the present example. In this example the distance measure is computed against a set of four predetermined standard images: the standard images were obtained by dividing a large test dataset of images into four different image types corresponding respectively to four different Cerb-B2 status indicators (as will be described later in more detail). The images of each image type were analysed to determine grand mean and relative range for each image as described above. A respective average grand mean M_i (i = 0, 1, 2 and 3) and a respective average relative range $skew_i$ were determined for the images of each of the four image types. As an alternative, it is also possible to select four good quality images of the relevant types by inspection from many images, and to determine M_i and $skew_i$ from them. The values M_i and $skew_i$ become the components of respective four-element vectors $\mathbf{M}$ and $\mathbf{skew}$, and are used in the following expression:

$$\text{Cerb-B2 indicator} = \min_i\left\{(M_i - grand\ mean)^2 + (skew_i - relative\ range)^2\right\} \quad (16)$$

where $\min_i$ is the value of i (i = 0, 1, 2 or 3) for which the expression within curved brackets { } on the right of Equation (16) is a minimum. For the vector $\mathbf{M}$, from the dataset the following elements were determined: $M_0 = 12.32$, $M_1 = 23.16$, $M_2 = 42.34$ and $M_3 = 87.35$); elements determined likewise for the vector $\mathbf{skew}$ were $skew_0 = 2.501$, $skew_1 = 1.85$, $skew_2 = 1.111$ and $skew_3 = 0.5394$. The value of i is returned as the indicator for the Cerb-B2 measurement process.

If a value of $i = 3$ is obtained in the Cerb-B2 measurement process, this is regarded as a strongly positive result: the patient from whom the original tissue samples were taken is regarded as highly suitable for treatment, currently with herceptin. A value of $i = 2$ is

weakly positive indicating doubtful suitability for treatment, and $I = 1$ or 0 is a negative result indicating unsuitability. This is tabulated below in Table 2.

TABLE 2

Cerb-B2 status	I Value
Strongly positive	3
Weakly positive	2
Negative	0, 1

Referring now to Figure 3, there is shown a flow diagram of the process 14 (see Figure 1) for measurement of vascularity. The process 14 is applied to two images each of x40 magnification compared to the histopathological slide from which they were taken. At 60 each image is transformed from red/green/blue (RGB) to a different image space hue/saturation/value (HSV). The RGB to HSV transformation is described by K. Jack in 'Video Demystified', 2nd ed., HighText Publications, San Diego, 1996. In practice value V (or brightness) is liable to vary due to staining and thickness variations across a slide, as well as possible vignetting by a camera lens used to produce the images. In consequence in this example the V component is ignored: it is not calculated, and emphasis is placed on the hue (or colour) and saturation values H and S. H and S are calculated for each pixel of the two RGB images as follows:

Let M = maximum of (R,G,B) (17)

Let m = minimum of (R,G,B) (18)

Then newr = (M − R)/(M − m) (19)

newg = (M − G)/(M − m) and (20)

newb = (M − B)/(M − m) (21)

This converts each colour of a pixel into the difference between its magnitude and that of the maximum of the three colour magnitudes of that pixel, this difference being divided by the difference between the maximum and minimum of (R,G,B).

Saturation (S) is set as follows:

if M equals zero, then S = 0 (22)

if M does not equal zero, then S = (M – m)/M (23)

The calculation for Hue (H) is as follows: from Equation (17) M must be equal to at least one of R, G and B:

if M equals zero, then H =180 (24)

If M equals R then H = 60(newb – newg) (25)

If M equals G then H = 60(2 + newr – newb) (26)

If M equals B then H = 60(4 + newg – newr) (27)

If H is greater than or equal 360 then H = H – 360 (28)

If H is less than 0 then H = H + 360 (29)

The Value V is not used in this example, but were it to be used it would be set to the maximum of (R,G,B).

The next step 62 is to apply colour segmentation to obtain a binary image. This segmentation is based on thresholding using the Hue and Saturation from the HSV colour space, and is shown in Table 3 below.

TABLE 3

Threshold Criterion	Binary Image Pixel Value
Pixel with both Hue H in the range 282 – 356 degrees (scale 0 to 360), and Saturation S in the range 0.2 to 0.24 (scale 0 to 1)	Set pixel to 1
Pixel with either Hue outside the range 282 – 356 degrees, and/or Saturation outside the range 0.2 – 0.24	Set pixel to 0

This produces a segmented binary image.

The next stage 64 is to apply connected component labelling (as defined previously) to the segmented binary image: this provides a binary image with regions of contiguous pixels equal to 1, the regions being uniquely labelled for further processing and their areas being determined. The labelled binary image is then spatially filtered to remove

small connected components (image regions with less than 10 pixels), which provides a reduced binary image.

The sum of the area of the remaining image regions in the reduced binary image is then determined at 66 from the results of connected component labelling, and this sum is then expressed as a percentage of the area of the whole image. This procedure is carried out for both of the original RGB images separately to provide two such percentage area values: the average of the two percentage area values is computed, and it represents an estimate of the percentage of the area of a tissue sample occupied by blood vessels – i.e. the sample vascularity.

As set out in Table 4 below, vascularity is determined to be high or low depending on whether or not it is equal to at least 31%.

TABLE 4

Description of vascularity	Range
High	31% – 100%
Low	0% – 30%

High vascularity corresponds to relatively fast tumour growth because tumour blood supply has been facilitated, and early treatment is indicated. Low vascularity corresponds to relatively slow tumour growth, and early treatment is less important.

Referring now to Figure 8, there is shown a flow diagram of a process 16 for the measurement of Oestrogen (ER) and Progesterone (PR). The process for measuring these two is the same except for their use of different immunohistochemical staining processes applied to tissue samples (slide colouring). The process will be described for the case of Oestrogen receptor assessment with diaminobenzidine as staining agent.

A digitised image of a histopathological slide or part of such a slide part can be characterised for its response to oestrogen, for any immunohistochemical stain used. A user can adapt the process to be described below by defining an appropriate reference colour, which must represent a sample with maximum intake of oestrogen. There are two main colours in the pixels of ER medical images, i.e. brown and blue. The same colours

appear in PR imagery, and so the image processing technique is the same. In this example the process determines brown pixels in an image that are similar to a reference brown colour defined by a user of the process.

Referring now to Figure 4, processing 16 to determine ER status will be outlined and then described in more detail later. It begins with a pre-processing stage 70 in which a K-means clustering algorithm is applied to a colour image using a Mahalanobis metric. This determines or cues image regions of interest for further processing by associating pixels into clusters on the basis of their having similar values of the Mahalanobis metric. At 72 the colour image is transformed into a chromaticity space which includes a location of a reference colour. Hue and saturation are calculated at 74 for pixels in clusters cued by K-means clustering. The number of brown stained pixels is computed at 76 by thresholding on the basis of hue and saturation. An ER status measurement is then derived at 78 from a combination of the fraction of stained pixels and average colour saturation.

The input for the ER preprocessing stage 70 consists of raw digital data files of a single colour histopathological image or tile. A triplet of image band values for each pixel represents the colour of that pixel in its red, green, and blue spectral components. These values in each of the three image bands are in the range [0...255], where [0,0,0] corresponds to black and [255,255,255] corresponds to white.

The K-means clustering algorithm 70 is applied to the digital colour image using clusters and the Mahalanobis distance metric. A cluster is a natural grouping of data having similar values of the relevant metric and the Mahalanobis distance metric is a measurement that gives an indication of degree of closeness of data items to a cluster centre. It is not essential to use four clusters or the Mahalanobis distance metric but these provide an adequate subdivision of the data and a convenient metric. The K-means algorithm is described by J. A. Hartigan and M. A. Wong, in a paper entitled 'A K-means clustering algorithm', Algorithm AS 136, in the Applied Statistics Journal, 1979. The Mahalanobis distance metric is described by F. Heijden, in 'Image Based Measurement Systems – object recognition and parameter estimation', John Wiley & Sons, 1994 and by R. Schalkoff , in 'Pattern Recognition – Statistical, Structural and Neural approaches', John Wiley & Sons Inc., 1992. The process comprises an initialisation step a) followed by computation of a covariance matrix at step b). This leads

to a likelihood calculation at step c), which effectively provides the distance of a pixel from a cluster centre. The procedure is as follows:

a) Initially, cluster centres are set at random locations and approximately equal numbers of pixels are arbitrarily assigned to each cluster for later readjustment.

b) For each cluster the following computations are carried out:

i) Compute elements of the kind σ_{ij}^k of a covariance matrix of the image bands indicating the degree of variation between intensities of different colours in pixels of each cluster :

$$\sigma_{ij}^k = \frac{1}{N^k}\sum_{l=1}^{N^k}\left(c_{il} - \mu_i^k\right)\left(c_{lj} - \mu_j^k\right) \qquad (30)$$

where: σ_{ij}^k is the ij^{th} element of the covariance matrix,

N^k is the number of pixels in cluster k,

c_{il} and c_{lj} are the values of pixel l in image bands i and j,

i, j take values 1, 2, 3, which represent the red, green and blue image bands respectively,

μ_i^k is the mean of all pixels in image band i belonging to cluster k, and

μ_j^k is the mean of all pixels in image band j belonging to cluster k.

ii) Calculate the determinant of the covariance matrix denoted as $\sum_{\det}^{k}$.

iii) Calculate the inverse of the covariance matrix denoted as $\sum_{inv}^{k}$.

c). With i denoting pixel number, each pixel $\vec{x}_i$ is now treated as a vector having three elements $x_{i,1}, x_{i,2}, x_{i,3}$ which are the red $(x_{i,1})$, green $(x_{i,2})$ and blue $(x_{i,3})$ pixel values: the red, green and blue image bands are therefore represented by second subscript indices 1, 2 and 3 respectively. With i ranging over all pixels in a cluster k, the likelihood $d^k(\vec{x}_i)$ of a pixel vector $\vec{x}_i$ not belonging to that cluster is computed from Equation (31) below:

$$d^k(\vec{x}_i) = \ln\left(\sqrt{\sum_{\det}^{k}}\right) + \frac{1}{2}\left[\left(\vec{x}_i - \vec{\mu}^k\right)^t \sum_{inv}^{k}\left(\vec{x}_i - \vec{\mu}^k\right)\right] \qquad (31)$$

where $\sum_{\text{red}}^{k}$ and $\sum_{\text{any}}^{k}$ are as defined above,

μ_t^k is the mean of all pixel vectors $\vec{x}_t$ in cluster k, and

t indicates the transpose of the difference vector $(\vec{x}_t - \vec{\mu}^k)$.

Equation (31) is re-evaluated for the same pixel vector $\vec{x}_t$ in all other clusters also. Pixel vector $\vec{x}_t$ has the highest likelihood of belonging to a cluster (denoted k_m) for which $d^k(\vec{x}_t)$ has a minimum value i.e. $\{d^{k_m}(\vec{x}_t)\}$; cluster k_m is then the most suitable to receive pixel $\vec{x}_t$; i.e. find:-

$$d^{k_m}(\vec{x}_t) \le d^k(\vec{x}_t) \quad \text{for all} \quad k \neq k_m \qquad (32)$$

Assign pixel $\vec{x}_t$ to cluster k_m

d). For each cluster k:

Store a record of which pixels belong to cluster k as an array X^k, update it with each pixel vector assigned to that cluster and update the number N^k of pixels in that cluster.

Calculate the cluster centre μ_j^k for each image band j = 1, 2 and 3

$$\mu_j^k = \frac{1}{N^k}\sum_{i=1}^{N^k} x_i^k \qquad (33)$$

Iterate steps b) to d) until convergence, i.e. when no more pixels change clusters or the number of iterations reaches a total of 20.

The first cluster (k = 1) now corresponds to cell nuclei and the corresponding pixel vectors are those which are cued as of interest for output and further processing.

Transformation of the image at 72 from red/green/blue (RGB) to hue/saturation/value (HSV) is as described by K. Jack in 'Video Demystified', 2nd ed., HighText Publications, San Diego, 1996. In practice value V (or brightness) is liable to vary due to staining and thickness variations across a slide, as well as possible vignetting by a camera lens used to produce the images. In consequence in this example the V component is ignored: it is

not calculated, and emphasis is placed on the hue (or colour) and saturation values H and S. H and S are calculated for each pixel of the RGB image as follows:

(a) Referring now also to Figures 5 to 8, each RGB image is transformed into a chromaticity space. Figure 5 shows an RGB cube 100 in which red, green and blue pixel values (expressed as R, G and B respectively) are normalised and represented as values in the range 0 to 1. These pixel values are represented on red, green and blue axes 102, 104 and 106 respectively. The chromaticity space is a plane 108 for which R+G+B = 1: it is triangular within the RGB cube 100 and passes through the points (1,0,0), (0,1,0) and (0,0,1).

(b) Figure 6 shows the axes 102, 104 and 106 and chromaticity space 108 looking broadly speaking along a diagonal of the RGB cube 100 from the point (1,1,1) (not shown) to the origin (0,0,0) now referenced O for convenience. The points (0,0,1), (0,1,0) and (1,0,0) in Figure 5 are now referenced J, K and L respectively. D is a midpoint of a straight line between J and L. Image pixel values from the input RGB image are projected on to the chromaticity space 108 and the resulting projections become data points for further processing. The projection calculation is as follows:

Red green and blue pixel chromaticity values r, g and b respectively are defined as:- $r = \frac{R}{R+G+B}$, $g = \frac{G}{R+G+B}$, and $b = \frac{B}{R+G+B}$ (34)

Perpendiculars from a point P in the chromaticity space 108 to the lines JK and LD meet the latter at E and G respectively. Perpendiculars from P and G to the plane JOK meet the latter at F and H respectively. Using Equations (34), the point P in the triangular chromaticity space 108 may then be defined by x and y co-ordinates shown in Figure 6 and given by:

$$x = DE = HF = \frac{g-r}{\sqrt{2}} \text{ and } y = PE = GD = b\sqrt{\frac{3}{2}} \quad (35)$$

(c) In Figure 7, the chromaticity space 108 is shown with x and y co-ordinate axes extending from an origin Q. A reference colour denoted by a point S in the drawing is now defined as that specified for this purpose by a clinician: it is the colour of that part of the image which is most positively stained (the most

intense colour on the part of the original slide from which the image was taken). The reference colour's RGB components are taken from the image and its x and y co-ordinates are computed using Equations (34) and (35); these co-ordinates are denoted as $(\tilde{x}, \tilde{y})$.

(d) In Figure 8, a polar co-ordinate system (r,θ) is now defined on the (R+G+B=1) plane 108. The co-ordinate system origin is the centre of gravity G of the triangle 109. A reference direction for θ = 0 is defined as the direction QS of the radius vector to the reference colour S in Figure 7. For any point such as P on the triangle defined as having co-ordinates (x, y) in the HSV colour space, hue H is defined as the angle φ between the radius vector (e.g. QP) to itself and the reference colour direction QS. This is computed from the following expressions for φ:

$$\sin\phi = \frac{\tilde{x}y - x\tilde{y}}{\sqrt{\tilde{x}^2 + \tilde{y}^2}\sqrt{x^2 + y^2}} \quad \text{and} \quad \cos\phi = \frac{x\tilde{x} + y\tilde{y}}{\sqrt{\tilde{x}^2 + \tilde{y}^2}\sqrt{x^2 + y^2}}$$

and the angle ϕ is defined to be $\sin^{-1}\dfrac{|\tilde{x}y - x\tilde{y}|}{\sqrt{\tilde{x}^2 + \tilde{y}^2}\sqrt{x^2 + y^2}}$

For convenience the definition of hue H is now altered somewhat to render all values positive and in the range 0 to π/2: the transformation of earlier values φ into a new version ψ is shown in Table 5 below:

TABLE 5

Condition	Magnitude of ψ (New Hue H)
$\sin\phi > 0$ and $\cos\phi > 0$	ϕ
$\sin\phi > 0$ and $\cos\phi < 0$	$\pi - \phi$
$\sin\phi < 0$ and $\cos\phi > 0$	$-\phi$
$\sin\phi < 0$ and $\cos\phi < 0$	$\phi - \pi$

A hue (H) threshold ψ_0 is set by a user or programmer of the procedure as being not more than $\pi/2$, a typical value which might be chosen being 80 degrees. Saturation S is defined to be

$$saturation = \frac{x\bar{x} + y\bar{y}}{\bar{x}^2 + \bar{y}^2}$$

Two values of saturation threshold S_0 are set according to whether or not image pixel saturation value S lies in the range 0.1 to 1.9: this is set out in Table 6 below:

TABLE 6

Saturation S	S_0
Either $S < 0.1$ or $S > 1.9$	0
$0.1 \leq S \leq 1.9$	0.9

At 106, the thresholds are used to count selectively the number N_b of pixels which are sufficiently brown (having a large enough value of saturation) having regard to the reference colour. All H and S pixel values in the image are assessed. The conditions to be satisfied by a pixel's hue and saturation values for it to be counted in the brown pixel number N_b are set out in Table 7 below.

TABLE 7

Condition	Action
For each pixel with both hue modulus $\|\psi\| < \psi_0$ and *saturation* $S > S_0$	Treat as a "saturated" pixel; increase count N_b of brown pixels by 1
For each pixel with $\|\psi\| \geq \psi_0$ and/or *saturation* $S \leq S_0$	Treat as an "unsaturated" pixel; leave N_b unchanged

The average saturation of the N_b saturated pixels determined in Table 7 is computed by adding all their saturation values S together and dividing the resulting sum by N_b. The maximum saturation value of the saturated pixels is then determined, and the average saturation is expressed as a percentage of this maximum: the average saturation is then accorded a score at 78 of 0, 1, 2 or 3 according respectively to whether this percentage is (a) ≤ 25%, (b) > 25% and ≤ 50%, (c) > 50% and ≤ 75% or (d) > 75% and ≤ 100%.

The fraction of saturated pixels - those stained sufficiently brown - is computed at 78 from the ratio N_b/N where N is the total number of pixels in the image. This fraction is then quantised to a score in the range 0 to 5 as set out in Table 8 below.

TABLE 8

N_b/N : Fraction of image pixels that are stained	Score
0.00	0
< 0.01	1
0.01 – 0.10	2
0.11 – 0.33	3
0.34 – 0.66	4
0.67 – 1.00	5

The two scores determined above, i.e. for average colour saturation and fraction of sufficiently brown pixels are now added together to give a measure in the range 0 to 8. The higher this number is, the more oestrogen (ER) positive the sample is.

Description of ER status (ER Score)	Range
Strongly positive	7 – 8
Positive	4 – 6
Weakly positive	2 – 3
Negative	0 – 1

Women with an ER score of 7 or 8 will respond favourably to hormonal treatment such as Tamoxifen; women with an ER score in the range 4 to 6 will have 50% of chance of responding to this treatment. Women scoring 2 or 3 will not respond very well, and those scoring 0 or 1 will not respond to hormonal treatment at all.

Images for ER and PR are indistinguishable visually and they are distinguished by the fact that they are produced using different stains. A PR score is therefore produced from stained slides in the same way as an ER score described above. The significance of progesterone receptor (PR) positivity in a breast carcinoma is less well understood than the equivalent for ER. In general, cancers that are ER positive will also be PR positive. However, carcinomas that are PR positive, but not ER positive, may have a worse prognosis.

The process steps described in the examples of all three inventions described herein are not all essential and alternatives may be provided. It is for example possible to omit a step of ignoring unsuitably small areas in or selecting areas for later processing, if the consequent increase in processing burden is acceptable. The above examples are intended to provide an enabling disclosure, not to limit the invention.

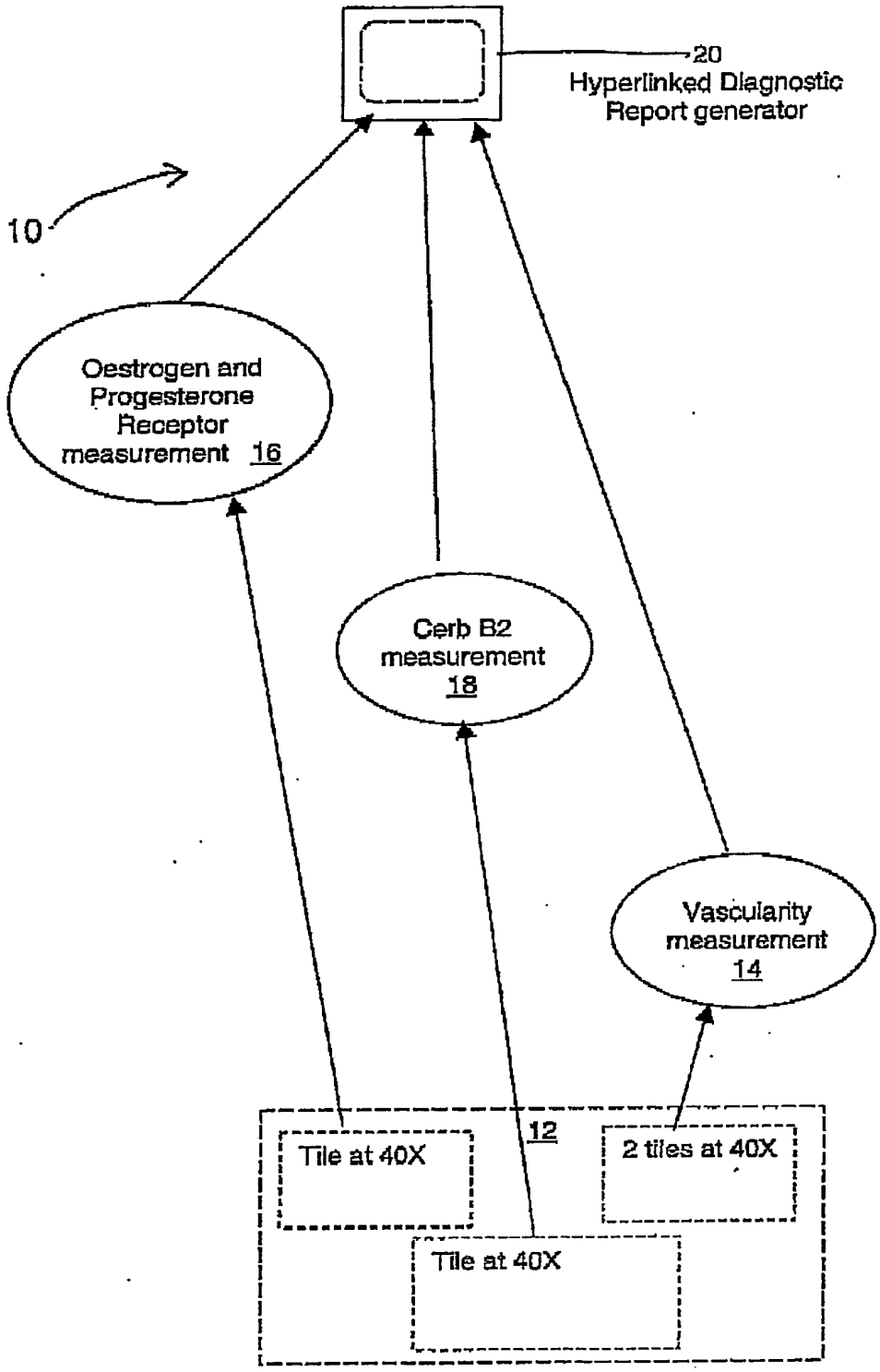

Figure 1

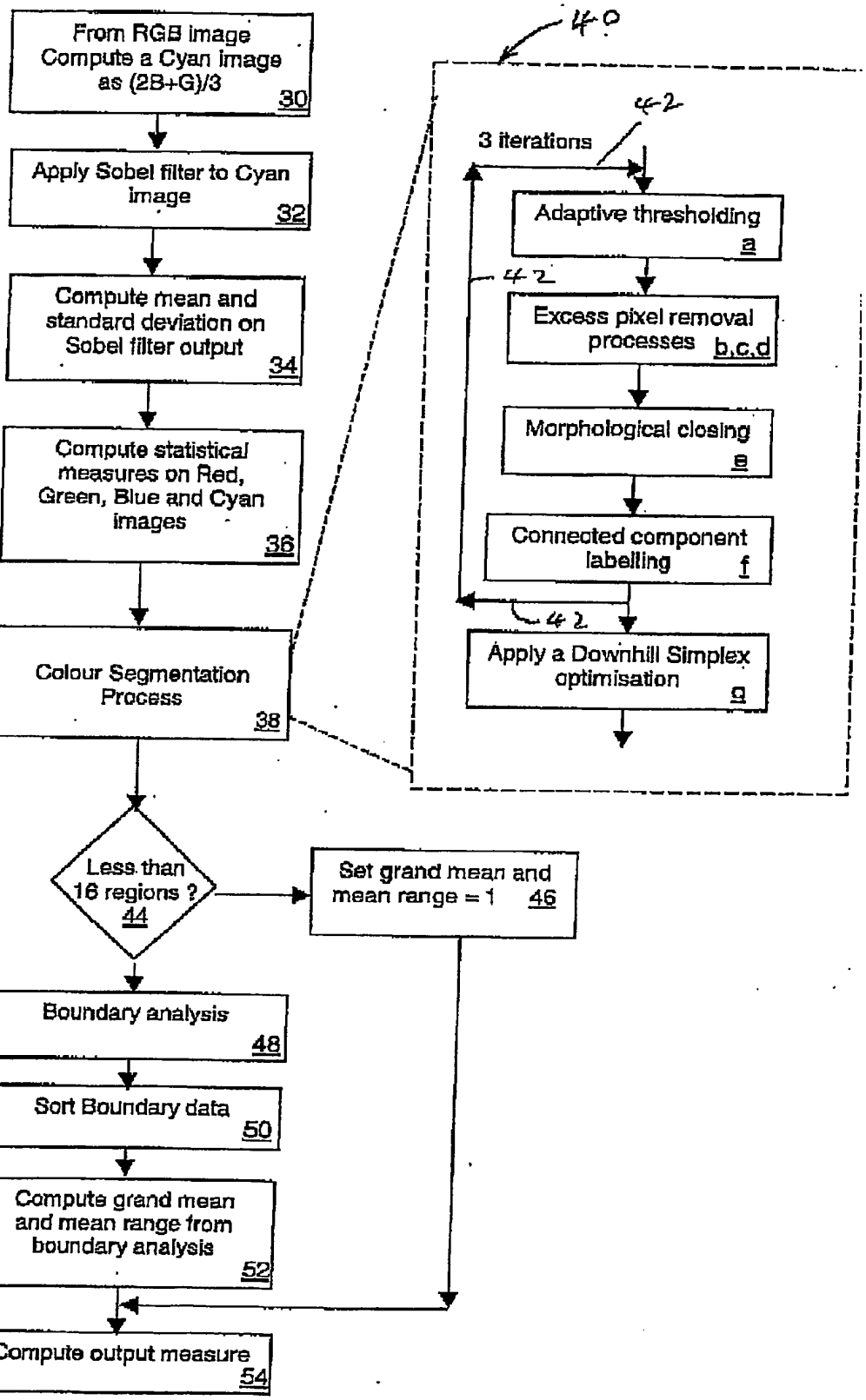

Figure 2

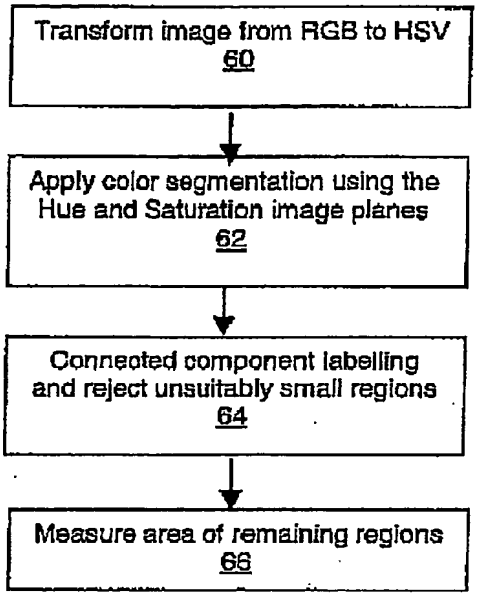

Figure 3 Vascularity process

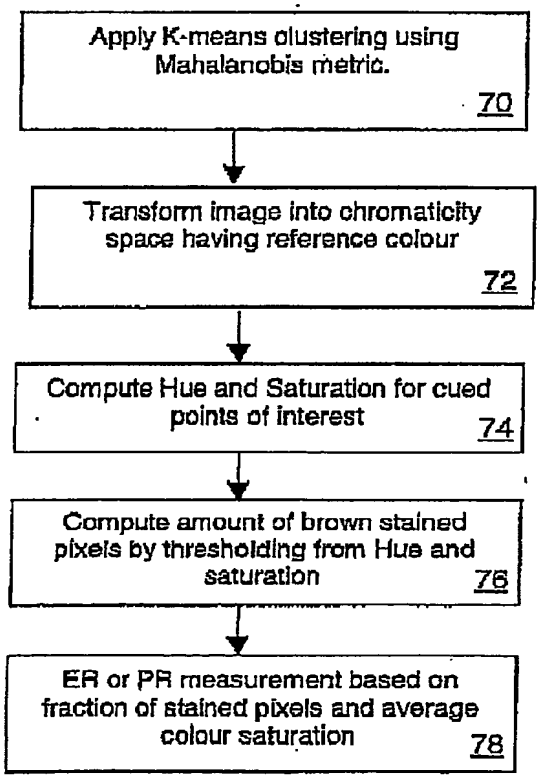

Figure 4 ER & PR measurement

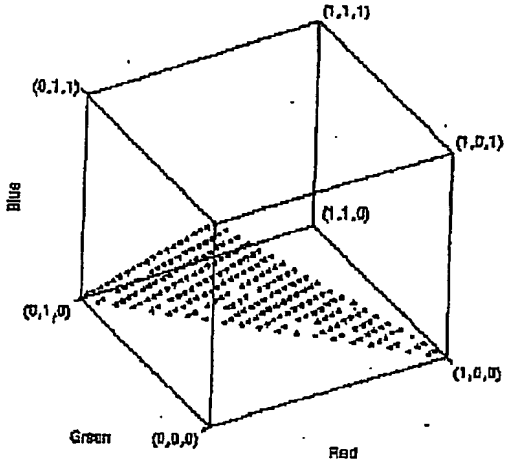

Figure 5 RGB colour cube

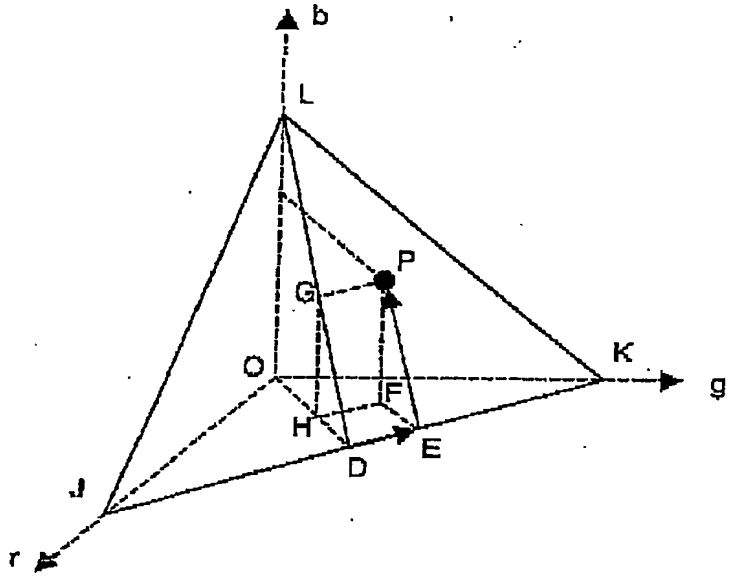

Figure 6 Chromaticity space

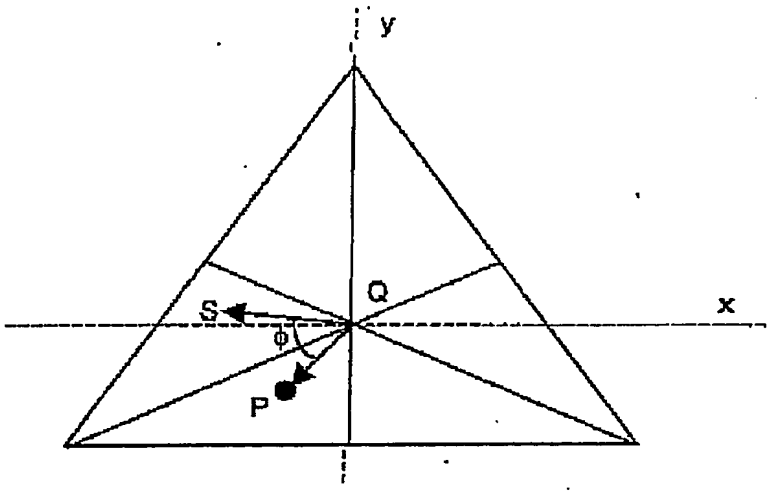

Figure 7 Chromaticity space reference system

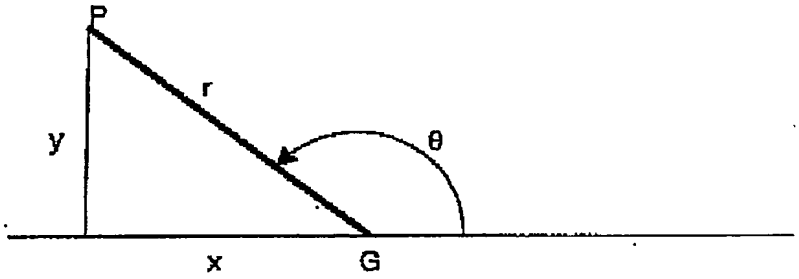

Figure 8 Polar coordinate system

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☑ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☑ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.